National Environmental Public Health Tracking Network
Downscaler PM 2.5 Metadata – County Level Data

Publication Date	01/11/2017
Background	The Downscaler $PM_{2.5}$ dataset provides the output from a Bayesian space-time downscaling fusion model called Downscaler (DS) that combines $PM_{2.5}$ monitoring data from the US EPA Air Quality System (AQS) repository of ambient air quality data (e.g., National Air Monitoring Stations/State and Local Air Monitoring Stations (NAMS/SLAMS)) and simulated $PM_{2.5}$ data from the deterministic prediction model, Models-3/Community Multiscale Air Quality (CMAQ). The files contain estimates for each of the 2010 U.S. counties within the contiguous U.S. for each day of the modeling year. The data are intended for use by professionals comparing air quality and health outcomes, through techniques such as case crossover analysis. Other uses may be developed at a later time.
Data Values	The dataset includes the following variables: STATEFIPS: State FIPS code COUNTYFIPS: County FIPS code YEAR: Year of estimate DATE: Date (day-month-year) of estimate PM25_MAX_PRED: Maximum estimated 24-hour average $PM_{2.5}$ concentration in $\mu g/m^3$ for the county PM25_MEAN_PRED: Mean estimated 24-hour average $PM_{2.5}$ concentration in $\mu g/m^3$ for the county PM25_MED_PRED: Median estimated 24-hour average $PM_{2.5}$ concentration in $\mu g/m^3$ for the county PM25_POP_PRED: Population-weighted estimated 24-hour average $PM_{2.5}$ concentration in µg/m3 for the county
Geographic Scale & Scope	All counties in the contiguous United States
Time Period	January 1, 2001 to December 31, 2014
Raw Data Processing	The air quality monitoring data from the NAMS/SLAMS network were downloaded from the Air Quality System (AQS) database. Only Federal Reference Method (FRM) samplers were included in the dataset. Data from all Pollutant Occurrence Codes (POC) were used. The data was downloaded covering January 1, 2001 through December 31, 2014. The CMAQ data was created from version 4.7.1 of the model using Carbon Bond Mechanism-05 (CB-05). The CMAQ data are daily 24-hour average $PM_{2.5}$ concentrations calculated on a 12 km x 12 km grid for the continental United States. The CMAQ emissions data are based on 2008 NEI version 2, with specific updates including data from regional planning organizations and year-specific data for some larger point sources, including continuous emissions monitoring data for NOx and SO2 sources. The onroad mobile source emissions were generated using MOVES 2010B, except for California, in which data provided by the California Air Resources Board was interpolated to each year. In addition, the meteorological data used are from the Weather Research and Forecasting Model (WRF) version 3.2 at 12 km simulation. The WRF simulation included the physics options of the Pleim-Xiu land surface model (LSM), Asymmetric Convective Model version 2 planetary boundary layer (PBL) scheme, Morrison double moment microphysics, Kain- Fritsch cumulus parameterization scheme and the RRTMG long-wave and shortwave radiation (LWR/SWR) scheme. The DS combines the actual monitoring

	data and the estimated $PM_{2.5}$ concentration surface (CMAQ) to predict $PM_{2.5}$ through space and time. It attempts to find an optimal linear relationship between CMAQ output and measurement data to predict new "measurements" at each spatial point in the area of interest. Fitted parameters are based on sampling from distributions (built into the code by the developers) rather than an objective function minimum, which allows calculation of a standard error associated with each prediction. Additional processing of the data was conducted to standardize variable names across all years of data and to expand FIPS variables into separate statefips and countyfips variables. Daily maximum, mean, median, and population-weighted values were calculated for each county based on census tract estimates and 2010 U.S. Census census tract-level population data.
Additional Information	Berrocal, V., Gelfand, A. E. and Holland, D. M. (2011). Space-time fusion under error in computer model output: an application to modeling air quality (http://onlinelibrary.wiley.com/doi/10.1111/j.1541-0420.2011.01725.x/abstract) Berrocal, V., Gelfand, A. E. and Holland, D. M. (2010). A bivariate space-time downscaler under space and time misalignment. The Annals of Applied Statistics 4, 1942-1975 Berrocal, V., Gelfand, A. E., and Holland, D. M. (2010). A spatio-temporal downscaler for output from numerical models. J. of Agricultural, Biological,and Environmental Statistics 15, 176-197